Integer-Valued Self-Exciting Threshold Autoregressive Processes

Magda Monteiro
Escola Superior de Tecnologia e Gestão de Águeda
Research Unit CIDMA
Universidade de Aveiro
Portugal

Manuel G. Scotto* and Isabel Pereira
Departamento de Matemática
Research Unit CIDMA
Universidade de Aveiro
Portugal

Abstract

In this paper we introduce a class of self-exciting threshold integer-valued autoregressive models driven by independent Poisson-distributed random variables. Basic probabilistic and statistical properties of this class of models are discussed. Moreover, parameter estimation is also addressed. Specifically, the methods of estimation under analysis are the least squares-type and likelihood-based ones. Their performance is compared through a simulation study.

Keywords: Count processes; Threshold models; Binomial thinning.

Mathematics Subject Classification: 62M10; 91B70; 60G10.

*Manuel G. Scotto, Departamento de Matemática, Universidade de Aveiro. Campo Universitário de Santiago, 3810-193 Aveiro, Portugal. Tel.: +351 234370670; fax: +351 234370670. Email: mscotto@ua.pt.

1 Introduction

In the analysis of stationary integer-valued time series the class of Poisson integer-valued autoregressive moving average models plays a central role. Such models, however, are unlikely to provide a sufficiently broad class capable of accurately capturing features often exhibited by time series of counts such as sudden burst of large values, volatility changes in time, high threshold exceedances appearing in clusters, and the so-called *piecewise* phenomenon. Addressing some of these issues Hall et al. (2010) introduced a general class of models specially devised for modeling non-negative integer-valued time series assuming low values with high probability but exhibiting, at the same time, sudden burst of large positive values. Doukhan et al. (2006) also gave a noticeable contribution by introducing a class of integer-valued bilinear models. Extensions of Doukhan and co-authors' work have been proposed by Drost et al. (2008). However, in the field of integer-valued time series data no efforts have been made so far to develop models for dealing with time series of counts exhibiting piecewise-type patterns. To the best of our knowledge only one contribution is known, namely Thyregod et al. (1999) who introduced a self-exciting threshold-based INAR (INteger-valued AutoRegressive) model to analyze tipping bucket rainfall measurements. In the work of Thyregod et al. (1999), however, a number of important issues related with the existence of the stationary marginal distribution of the process, the existence of moments, and the asymptotic distribution of the maximum likelihood estimators are left as open questions. This paper aims to give a contribution towards this direction.

Since their introduction by Tong (1977) much attention has been given to threshold models partially because of their wide applicability to economy and finance (Boero and Marrocu, 2004; Pai and Pedersen 1999; Potter, 1995), hydrological (Fu et al. 2004), ocean engineering (Scotto and Guedes Soares, 2000), electricity markets (Amaral et al., 2008) and physical phenomena (Tong, 1990). Among the more successful threshold models we mention the Self-Exciting Threshold AutoRegressive Moving Average (in short SETARMA) model (Tong, 1983). The SETARMA model of order $(k; p_1, \ldots, p_k; q_1, \ldots, q_k)$ takes the form

$$X_t = \sum_{i=1}^{k} \left[\phi_0^{(i)} + \sum_{j=1}^{p_i} \phi_j^{(i)} X_{t-j} + Z_t - \sum_{r=1}^{q_i} \psi_r^{(i)} Z_{t-r} \right] I(X_{t-d} \in R_i),\ t \in \mathbb{Z}, \tag{1}$$

where $(Z_t)_{t\in\mathbb{Z}}$ is a sequence of independent and identically distributed (i.i.d.) random variables (r.v.'s), $R_i := [r_{i-1}, r_i)$ forms a partition of the real line such that $-\infty = r_0 < r_1 < \cdots < r_k = +\infty$, being r_i the threshold values, d represents the threshold delay, p_i and q_i are non-negative integers referred to the AR and MA orders, respectively, and $\phi_j^{(i)}$ and $\psi_r^{(i)}$ are unknown parameters, for $j = 1, \ldots, p_i$ and $r = 1, \ldots, q_i$. Finally, $I(\cdot)$ is a Bernoulli random process. Note that the SETARMA model is characterized by a piecewise linear structure which follows a conventional linear ARMA model in each regime R_i with $\bigcup_{i=1}^{k} R_i = \mathbb{R}$. It is worth noting that this model is piecewise linear in the space of the threshold variable rather than in time. The model in (1) is appealing from a physical perspective as many physical systems are state dependent in the sense that the nature of their future evolution is dependent on their current state. A number of such examples are discussed by Tong (1990).

It is worth to mention that all references given in the previous paragraph deal with the case of conventional (*id est*, continuous-valued) threshold models. In contrast, however, the analysis of integer-valued threshold models has not received much attention in the literature. Motivation to include discrete data models comes from the need to account for the discrete nature of certain data sets, often counts of events, objects or individuals. The analysis of time series of counts has become an important area of research in the last two decades partially because of its wide applicability to social science (McCabe and Martin, 2005), queueing systems (Ahn et al., 2000), experimental biology (Zhou and Basawa, 2005), environmental processes (Thyregod et al., 1999), economy (Brännäs and Quoreshi, 2010), statistical control processes (Weiß, 2008c; Lambert and Liu, 2006), telecommunications (Weiß, 2008a), optimal alarm systems (Monteiro et al., 2008), and in the biopharmaceutical industry (Alosh, 2009). We refer to McKenzie (2003) for an overview of the early work in this area and to Jung and Tremayne (2006, 2010) and Weiß (2008b) for recent developments.

In this paper, we investigate basic probabilistic and statistical properties of the self-exciting threshold integer-valued autoregressive model of order one with two regimes (hereafter referred to as SETINAR$(2,1)$) defined by the recursive equation

$$X_t = \phi_t \circ X_{t-1} + Z_t, \; t \in \mathbb{Z}, \tag{2}$$

with $\phi_t := \alpha_1 I_{t-1,1} + \alpha_2 I_{t-1,2}$, where the *thinning* operator $\circ$ is defined as

$$\phi_t \circ X_{t-1} \overset{d}{=} I_{t-1,1} \sum_{i=1}^{X_{t-1}} U_{i,t}(\alpha_1) + I_{t-1,2} \sum_{i=1}^{X_{t-1}} U_{i,t}(\alpha_2),$$

being $(U_{i,t}(\alpha_1))$ and $(U_{i,t}(\alpha_2))$, for $i = 1, 2, \ldots$, i.i.d. sequences of Bernoulli random variables with success probabilities $P(U_{i,t}(\alpha_1) = 1) = \alpha_1 \in (0,1)$ and $P(U_{i,t}(\alpha_2) = 1) = \alpha_2 \in (0,1)$, respectively, which for each t both are independent of X_s for $s \leq t-1$. Moreover

$$I_{t-1,1} := \begin{cases} 1 & X_{t-1} \leq R \\ 0 & X_{t-1} > R \end{cases},$$

where R is the threshold level and $I_{t-1,2} = 1 - I_{t-1,1}$. Furthermore, throughout the paper we shall assume that R is *known* and that $(Z_t)_{t\in\mathbb{Z}}$ constitutes an i.i.d. sequence of Poisson-distributed random variables with mean λ, which for each t, Z_t is assumed to be independent of X_{t-1}, ϕ_t and $\phi_t \circ X_{t-1}$. Note that the operator $\circ$ incorporates the discrete nature of the variates and acts as the analogue of the standard multiplication used in the continuous-valued processes. This operator was first introduced by Steutel and van Harn (1979), to adapt the terms of self-decomposability and stability for integer-valued time series. Various modifications of this thinning operator have been proposed to make the integer-valued models based on thinning more flexible for practical purposes; see Weiß(2008b) for further details.

(Figure 1 about here)

Figure 1 shows two simulated sample paths from the SETINAR(2,1) model with (a): $\lambda = 7, \alpha_1 = 0.2, \alpha_2 = 0.65, R = 14$ and (b): $\lambda = 3$, and $\alpha_1 = 0.5, \alpha_2 = 0.65, R = 7$, respectively. The sample path in Figure 1(b) tends to move between regimes quite often reflecting the fact that when α_1 and α_2 are close from each other, it becomes more difficult to distinguish between the two regimes. This is in contrast with the sample path displayed is

Figure 1(a) in which the presence of two regimes becomes more obvious. This feature is also visible in Figure 2 in which the directed scatter diagrams for x_t against x_{t-1} and x_t against x_{t-2} are displayed for the two simulated sample paths generated by the SETINAR(2,1) models (a) and (b). Note that in Figure 2(a) there are only a few lines linking the two regimes whereas in Figure 2(b) is more difficult to distinguish between them.

(Figure 2 about here)

The rest of the paper is organized as follows: in Section 2, we demonstrate the existence of a strictly stationary SETINAR$(2,1)$-process satisfying (2). Expressions for the mean and variance are also given. Furthermore, we derive a set of equations from which the autocorrelation function can be obtained. Parameter estimation is covered in Section 3. In Section 4 the results are illustrated through a simulation study. Finally, some concluding remarks are given in Section 5.

2 Basic properties of the SETINAR$(2,1)$ model

Let X_t be the process defined in (2). We first prove that there exists a strictly stationary SETINAR$(2,1)$-process satisfying (2).

Proposition 2.1. *The process $(X_t)_{t\in\mathbb{Z}}$ is an irreducible, aperiodic and positive recurrent (and hence ergodic) Markov chain. Thus there exists a strictly stationary process satisfying (2).*

Proof. It is easy to see that X_t is a Markov chain with state space $\mathbb{N}_0$ with the following transition probabilities

$$\begin{aligned} P(X_t=j|X_{t-1}=i) &= \sum_{m=0}^{\min(i,j)} C_m^i \left(I_{t-1,1}\alpha_1^m(1-\alpha_1)^{i-m}+I_{t-1,2}\alpha_2^m(1-\alpha_2)^{i-m}\right)e^{-\lambda}\frac{\lambda^{j-m}}{(j-m)!} \\ &= p(i,j,\alpha_1,\lambda)I_{t-1,1}+p(i,j,\alpha_2,\lambda)I_{t-1,2} \\ &= p(i,j,\alpha_1 I_{t-1,1}+\alpha_2 I_{t-1,2},\lambda), \end{aligned}$$

where

$$p(i,j,\alpha_k,\lambda) := \sum_{m=0}^{\min(i,j)} C_m^i \alpha_k^m(1-\alpha_k)^{i-m}e^{-\lambda}\frac{\lambda^{j-m}}{(j-m)!} > 0,\ k=1,2. \tag{3}$$

From the expression above it follows that the chain is irreducible and aperiodic. Furthermore, to show that X_t is positive recurrent it is sufficient to prove that $\sum_{t=1}^{+\infty} P^t(0,0) = +\infty$ (since X_t is irreducible) with $P^t(x,y) := P(X_t = y|X_0 = x)$. Note that by iterating equation (2) it follows that

$$X_t = \phi_t \circ \phi_{t-1} \circ \cdots \circ \phi_1 \circ X_0 + \sum_{i=1}^{t-1} \phi_{t-1} \circ \phi_{t-2} \circ \cdots \circ \phi_{t-i} \circ Z_{t-i} + Z_t,$$

which allow us to write

$$\begin{aligned}
P^t(0,0) &= P\left(\sum_{i=1}^{t-1} \phi_{t-1} \circ \phi_{t-2} \circ \cdots \circ \phi_{t-i} \circ Z_{t-i} + Z_t = 0|X_0 = 0\right) \\
&= P(Z_t = 0, \phi_{t-1} \circ Z_{t-1} = 0, \ldots, \phi_{t-1} \circ \phi_{t-2} \circ \cdots \circ \phi_1 \circ Z_1 = 0|X_0 = 0) \\
&= \sum_{i_2=1}^{2} \cdots \sum_{i_{t-2}=1}^{2} \sum_{i_{t-1}=1}^{2} P(\phi_2 = \alpha_{i_2}, \ldots, \phi_{t-2} = \alpha_{i_{t-2}}, \phi_{t-1} = \alpha_{i_{t-1}}|X_0 = 0) \times \\
&\quad \times P(Z_t = 0, \alpha_{i_{t-1}} \circ Z_{t-1} = 0, \ldots, \alpha_{i_{t-1}} \circ \alpha_{i_{t-2}} \circ \cdots \circ \alpha_{i_2} \circ Z_1 = 0|X_0 = 0) \\
&= \sum_{i_2=1}^{2} \cdots \sum_{i_{t-2}=1}^{2} \sum_{i_{t-1}=1}^{2} P(\phi_2 = \alpha_{i_2}, \ldots, \phi_{t-2} = \alpha_{i_{t-2}}, \phi_{t-1} = \alpha_{i_{t-1}}|X_0 = 0) \times \\
&\quad \times e^{-\lambda(1+\alpha_{i_{t-1}}+\alpha_{i_{t-1}}\alpha_{i_{t-2}}+\cdots+\alpha_{i_{t-1}}\alpha_{i_{t-2}}\cdots\alpha_{i_2})}.
\end{aligned}$$

Note that the last expression implies that

$$e^{-\lambda \frac{1-\max(\alpha_1,\alpha_2)^t}{1-\max(\alpha_1,\alpha_2)}} \leq P^t(0,0) \leq e^{-\lambda \frac{1-\min(\alpha_1,\alpha_2)^t}{1-\min(\alpha_1,\alpha_2)}}.$$

Since $P^t(0,0) > 0$ it follows easily that $\sum_{i=1}^{+\infty} P^t(0,0) = +\infty$, by using the comparison criterion for series convergence . This proves that X_t is a positive recurrent Markov chain and hence ergodic which ensures the existence of a strictly stationary distribution of (2). $\square$

Remark 2.1. *As in the conventional case, it is generally difficult to obtain an explicit analytic formula for the stationary marginal distribution of the SETINAR process. In a companion paper, this issue will be treated and discussed in detail.*

The next lemma ensures that the first three moments exist. This lemma will be useful in proving some asymptotic properties of the conditional least squares estimators.

Lemma 2.1. *Let X_t be the process defined by the equation in (2). Then $E(X_t^k) \leq C < \infty$, for some constant $C > 0$, for $k = 1, 2, 3$.*

Proof. Let the chain X_t start in 0 at $t = 0$. From Proposition 2.1 and the theory of Markov chains it follows that $X_t \xrightarrow{d} Z$ where Z follows the stationary (marginal) distribution.

For $k = 1$, the aim is to prove that

$$E(X_t) \leq \alpha_{max}^t E(X_0) + \mu_z \sum_{i=0}^{t-1} \alpha_{max}^i \tag{4}$$

with $\alpha_{max} := \max(\alpha_1, \alpha_2)$, for any value of t. It is easy to check that the above inequality holds for $E(X_1)$. Furthermore, assume that is true for $E(X_{t-1})$, then

$$\begin{aligned} E(X_t) &\leq \alpha_{max} E(X_{t-1}) + \mu_z \\ &\leq \alpha_{max} \left(\alpha_{max}^{t-1} E(X_0) + \mu_z \sum_{i=0}^{t-2} \alpha_{max}^i \right) + \mu_z \\ &\leq \alpha_{max}^t E(X_0) + \mu_z \sum_{i=0}^{t-1} \alpha_{max}^i . \end{aligned}$$

For $k = 2$, it follows that

$$\begin{aligned} E(X_t^2) &\leq \alpha_{max}^{2t} E(X_0^2) + \left(\frac{1}{4} \alpha_{max}^{t-1} E(X_0) + 2\mu_z E(X_0) \alpha_{max}^t \right) \sum_{i=0}^{t-1} \alpha_{max}^i + \\ &+ \left(\sigma_Z^2 + \mu_z^2 + \frac{\mu_z}{4(1-\alpha_{max})} + 2 \frac{\mu_z^2 \alpha_{max}}{1-\alpha_{max}} \right) \sum_{i=0}^{t-1} \alpha_{max}^{2i} . \end{aligned} \tag{5}$$

Similarly, for $k = 3$

$$\begin{aligned} E(X_t^3) &\leq \alpha_{max}^{3t} E(X_0^3) + \left(\frac{3}{2} + 3\alpha_{max}^2 \mu_z \right) \times \alpha_{max}^{2t-2} E(X_0^2) \sum_{i=0}^{t-1} \alpha_{max}^i + \\ &+ \left\{ \left(\frac{3}{2} + 3\alpha_{max}^2 \mu_z \right) const_1 + \frac{1}{2} + \frac{3}{4}\mu_z + 3\alpha_{max}(\sigma_Z^2 + \mu_z) \right\} E(X_0) \alpha_{max}^{t-1} \times \\ &\times \sum_{i=0}^{t-1} \alpha_{max}^{2i} + \left\{ \left(\frac{3}{2} + 3\alpha_{max}^2 \mu_z \right) const_2 + \right. \\ &+ \left. \left(\frac{1}{2} + \frac{3}{4}\mu_z + 3\alpha_{max}(\sigma_Z^2 + \mu_z) \right) \frac{\mu_z}{1-\alpha_{max}} + E(Z_t^3) \right\} \sum_{i=0}^{t-1} \alpha_{max}^{3i} . \end{aligned} \tag{6}$$

In view of the fact that the chain starts at 0, by (4), (5) and (6) it follows that $E(X_t^k) < \infty$ for $k = 1, 2, 3$. Now, from the Portmanteau lemma (see e.g. Billingsley, 1979, Theorem 29.1,

p. 329) for convergence in distribution the result follows

$$E(Z^k) \le \lim_{t\to\infty} E(X_t^k) < \infty.$$

$\square$

Remark 2.2. *Note that* $E(X_t^k) < \infty$ *implies* $m_k^{(1)} := E(X_t^k | I_{t,1} = 1) < \infty$ *and* $m_k^{(2)} := E(X_t^k | I_{t,2} = 1) < \infty$ *for* $k = 1, 2, 3$.

Now we are prepared to obtain the mean and the autocovariance function of the process. For simplicity in notation we define $p := P(X_t \le R)$, $u_1 := E(X_t | X_t \le R)$, $u_2 := E(X_t | X_t > R)$, $\sigma_1^2 := V(X_t | X_t \le R)$, $\sigma_2^2 := V(X_t | X_t > R)$, and $\gamma_k^{(1)} := Cov(X_t, X_{t+k} | X_{t+k} \le R)$ and $\gamma_k^{(2)} := Cov(X_t, X_{t+k} | X_{t+k} > R)$.

Lemma 2.2. *The expectation of* X_t *is given by*

$$u := E(X_t) = p\alpha_1 u_1 + (1-p)\alpha_2 u_2 + \lambda.$$

Moreover, the variance of X_t *takes the form*

$$\begin{aligned} \sigma^2 := V(X_t) &= p(\alpha_1^2\sigma_1^2 + \alpha_1(1-\alpha_1)u_1) + (1-p)(\alpha_2^2\sigma_2^2 + \alpha_2(1-\alpha_2)u_2) + \lambda + \\ &+ p(1-p)(\alpha_1 u_1 - \alpha_2 u_2)^2. \end{aligned}$$

Finally, the autocovariance function $\gamma(k) := Cov(X_t, X_{t+k})$ *is given by*

$$\gamma(k) = \begin{cases} p\alpha_1\sigma_1^2 + (1-p)\alpha_2\sigma_2^2 + p\alpha_1 u_1(u_1 - u) + (1-p)\alpha_2 u_2(u_2 - u) & k = 1 \\ p\alpha_1\gamma_{k-1}^{(1)} + (1-p)\alpha_2\gamma_{k-1}^{(2)} + p\alpha_1\mu_1(E(X_t | X_{t+k-1} \le R) - u) & k \ne 1 \\ \quad +(1-p)\alpha_2 u_2(E(X_t | X_{t+k-1} > R) - u) & \end{cases}.$$

Proof.

$$\begin{aligned} E(X_t) &= E(\phi_{t-1} \circ X_{t-1}) + \lambda \\ &= E(I_{t-1,1} \sum_{i=1}^{X_{t-1}} U_{i,t}(\alpha_1)) + E(I_{t-1,2} \sum_{i=1}^{X_{t-1}} U_{i,t}(\alpha_2)) + \lambda \\ &= p\alpha_1 E(X_{t-1} | X_{t-1} \le r) + (1-p)\alpha_2 E(X_{t-1} | X_{t-1} > r) + \lambda \\ &= p\alpha_1 u_1 + (1-p)\alpha_2 u_2 + \lambda. \end{aligned}$$

Moreover

$$
\begin{aligned}
V(X_t) &= V(\phi_{t-1} \circ X_{t-1}) + \lambda \\
&= V(I_{t-1,1} \sum_{i=1}^{X_{t-1}} U_{i,t}(\alpha_1) + I_{t-1,2} \sum_{i=1}^{X_{t-1}} U_{i,t}(\alpha_2)) + \lambda \\
&= V(I_{t-1,1} \sum_{i=1}^{X_{t-1}} U_{i,t}(\alpha_1)) + V(I_{t-1,2} \sum_{i=1}^{X_{t-1}} U_{i,t}(\alpha_2)) + \\
&+ 2Cov\left(I_{t-1,1} \sum_{i=1}^{X_{t-1}} U_{i,t}(\alpha_1), I_{t-1,2} \sum_{i=1}^{X_{t-1}} U_{i,t}(\alpha_2) \right) + \lambda \\
&= I + II + III + \lambda. \qquad (7)
\end{aligned}
$$

The first term on the right-hand side of (7) is

$$
\begin{aligned}
I &= V(E[I_{t-1,1} \sum_{i=1}^{X_{t-1}} U_{i,t}(\alpha_1)|X_{t-1}]) + E(V[I_{t-1,1} \sum_{i=1}^{X_{t-1}} U_{i,t}(\alpha_1)|X_{t-1}]) \\
&= V(I_{t-1,1}\alpha_1 X_{t-1}) + E(I_{t-1,1}\alpha_1(1-\alpha_1)X_{t-1}) \\
&= \alpha_1^2 V(I_{t-1,1}X_{t-1}) + p\alpha_1(1-\alpha_1)u_1 \\
&= \alpha_1^2 E(I_{t-1,1}X_{t-1}^2) - \alpha_1^2 p^2 u_1^2 + p\alpha_1(1-\alpha_1)u_1 \\
&= \alpha_1^2 p(\sigma_1^2 + u_1^2) - \alpha_1^2 p^2 u_1^2 + p\alpha_1(1-\alpha_1)u_1 \\
&= p(\alpha_1^2\sigma_1^2 + \alpha_1(1-\alpha_1)u_1) + p(1-p)\alpha_1^2 u_1^2. \qquad (8)
\end{aligned}
$$

By the same arguments as above, it follows that

$$
II = (1-p)(\alpha_2^2\sigma_2^2 + \alpha_2(1-\alpha_2)u_2) + p(1-p)\alpha_2^2 u_2^2. \qquad (9)
$$

Finally, III takes the form

$$
III = -2\prod_{j=1}^{2} E\left(I_{t-1,j} \sum_{i=1}^{X_{t-1}} U_{i,t}(\alpha_j) \right) = -2p(1-p)\prod_{j=1}^{2} \alpha_j u_j. \qquad (10)
$$

Thus, the second statement in Lemma 2.2 follows by replacing (8), (9), and (10) in (7). The autocovariance function follows by similar arguments after some tedious calculations. We skip the details.

$\square$

3 Parameters estimation

Let $(X_1, \ldots, X_n)$ be a sequence of r.v's satisfying (2) being $\boldsymbol{\theta} := (\theta_1, \theta_2, \theta_3) \equiv (\alpha_1, \alpha_2, \lambda)$ the vector of unknown parameters. The methods of estimation under analysis in this section are the least squares and the conditional maximum likelihood. Recall from the introduction that R is assumed to be known.

3.1 Conditional Least Squares Estimators (CLS)

The CLS-estimators $\hat{\boldsymbol{\theta}}_{CLS} := (\hat{\alpha}_{1,CLS}, \hat{\alpha}_{2,CLS}, \hat{\lambda}_{CLS})$ of $\boldsymbol{\theta}$ are obtained by minimizing the expression

$$Q(\boldsymbol{\theta}) := \sum_{t=2}^{n} \left(X_t - g\left(\boldsymbol{\theta}, X_{t-1}\right)\right)^2 = \sum_{t=2}^{n} U_t^2(\boldsymbol{\theta})$$

where

$$g\left(\boldsymbol{\theta}, X_{t-1}\right) := \alpha_1 X_{t-1} I_{t-1,1} + \alpha_2 X_{t-1} I_{t-1,2} + \lambda,$$

yielding the system

$$\begin{bmatrix} \sum_{t=2}^{n} X_{t-1}^2 I_{t-1,1} & 0 & \sum_{t=2}^{n} X_{t-1} I_{t-1,1} \\ 0 & \sum_{t=2}^{n} X_{t-1}^2 I_{t-1,2} & \sum_{t=2}^{n} X_{t-1} I_{t-1,2} \\ \sum_{t=2}^{n} X_{t-1} I_{t-1,1} & \sum_{t=2}^{n} X_{t-1} I_{t-1,2} & n-1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \lambda \end{bmatrix} = \begin{bmatrix} \sum_{t=2}^{n} X_t X_{t-1} I_{t-1,1} \\ \sum_{t=2}^{n} X_t X_{t-1} I_{t-1,2} \\ \sum_{t=2}^{n} X_t \end{bmatrix}.$$

The following result establishes the asymptotic distribution of the CLS-estimators.

Theorem 3.1. *The CLS-estimators are strongly consistent and asymptotically normal, i.e.,*

$$n^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}_{CLS} - \boldsymbol{\theta}) \xrightarrow{d} N(0, V^{-1} W V^{-1}),$$

where V and W are square matrices of order 3, with elements

$$V_{ij} := E\left[\frac{\partial}{\partial \theta_i} g(\boldsymbol{\theta}, X_{t-1}) \frac{\partial}{\partial \theta_j} g(\boldsymbol{\theta}, X_{t-1})\right]$$

and

$$W_{ij} := E\left[U_t^2(\boldsymbol{\theta}) \frac{\partial}{\partial \theta_i} g(\boldsymbol{\theta}, X_{t-1}) \frac{\partial}{\partial \theta_j} g(\boldsymbol{\theta}, X_{t-1})\right],$$

respectively.

Proof. Consistency and asymptotic normality can be easily proved by using the results in Klimko and Nelson (1978, Section 3). First, it is easy to check that $g(\boldsymbol{\theta}, X_{t-1})$, $\frac{\partial g(\boldsymbol{\theta}, X_{t-1})}{\partial \theta_i}$, $\frac{\partial^2 g(\boldsymbol{\theta}, X_{t-1})}{\partial \theta_i \partial \theta_j}$ and $\frac{\partial^3 g(\boldsymbol{\theta}, X_{t-1})}{\partial \theta_i \partial \theta_j \partial \theta_k}$ for $i, j, k \in \{1, 2, 3\}$ satisfy all the regularity conditions in Klimko and Nelson (1978, p. 634). Thus, Theorem 3.1 in Klimko and Nelson (1978) lead us to conclude that the CLS-estimators are strongly consistent. Moreover, in proving asymptotic normality we have to check first that the following conditions holds:

(A) $E\left(X_t | X_{t-1}, X_{t-2}, \ldots, X_0\right) = E\left(X_t | X_{t-1}\right)$, $t \geq 1$ a.e.;

(B) $E\left(U_t^2(\boldsymbol{\theta}) \left| \frac{\partial}{\partial \theta_i} g(\boldsymbol{\theta}, F_{t-1}) \frac{\partial}{\partial \theta_j} g(\boldsymbol{\theta}, F_{t-1}) \right| \right) < \infty$, $i, j = 1, 2, 3$, where $U_t = X_t - g(\boldsymbol{\theta}, F_{t-1})$, $F_{t-1} = \sigma\left(X_s, s \leq t-1\right)$ and $g(\boldsymbol{\theta}, F_{t-1}) \equiv g\left(\boldsymbol{\theta}, X_{t-1}\right)$;

(C) V is non-singular.

Condition **(A)** is satisfied since X_t is a first-order Markov chain. In order to prove condition **(B)** we check that the W_{ij}'s for $i, j = 1, 2, 3$ are all finite.

$$\begin{aligned}
W_{1,1} = E&\left(U_t^2(\boldsymbol{\theta}) \left(\frac{\partial}{\partial \alpha_1} g(\boldsymbol{\theta}, F_{t-1})\right)^2\right) \\
&= E\left(U_t^2(\boldsymbol{\theta}) X_{t-1}^2 I_{t-1,1}\right) \\
&= E\left(X_{t-1}^2 I_{t-1,1} E\left(U_t^2(\boldsymbol{\theta}) | X_{t-1}\right)\right) \\
&= E\left(X_{t-1}^2 I_{t-1,1} V\left(X_t | X_{t-1}\right)\right) \\
&= E\left(X_{t-1}^2 I_{t-1,1} (\alpha_1(1-\alpha_1) X_{t-1} I_{t-1,1} + \alpha_2(1-\alpha_2) X_{t-1} I_{t-1,1} + \lambda)\right) \\
&= \alpha_1(1-\alpha_1) E(X_{t-1}^3 I_{t-1,1}) + \lambda E(X_{t-1}^2 I_{t-1,1}) \\
&= p\alpha_1(1-\alpha_1) m_3^{(1)} + p\lambda m_2^{(1)} \\
&< \infty \text{ (by Remark 2.2).}
\end{aligned}$$

Using the same arguments for α_2 we obtain

$$\begin{aligned}
W_{2,2} &= E\left(U_t^2(\boldsymbol{\theta}) \left(\frac{\partial}{\partial \alpha_2} g(\boldsymbol{\theta}, F_{t-1})\right)^2\right) \\
&= (1-p)\alpha_2(1-\alpha_2) m_3^{(2)} + (1-p)\lambda m_2^{(2)} \\
&< \infty \text{ (by Remark 2.2).}
\end{aligned}$$

Considering now $i = 3$ and $j = 3$, we have

$$
\begin{aligned}
W_{3,3} = E\left(U_t^2(\boldsymbol{\theta})\left(\frac{\partial}{\partial\lambda}g(\boldsymbol{\theta}, F_{t-1})\right)^2\right) \\
&= E\left(V\left(X_t|X_{t-1}\right)\right) \\
&= E\left((\alpha_1(1-\alpha_1)X_{t-1}I_{t-1,1} + \alpha_2(1-\alpha_2)X_{t-1}I_{t-1,1} + \lambda)\right) \\
&= \alpha_1(1-\alpha_1)E(X_{t-1}I_{t-1,1}) + \alpha_2(1-\alpha_2)E(X_{t-1}I_{t-1,2}) + \lambda \\
&= p\alpha_1(1-\alpha_1)E(X_{t-1}|X_{t-1} \leq R) + (1-p)\alpha_2(1-\alpha_2)E(X_{t-1}|X_{t-1} > R) + \lambda \\
&= u \text{ (by Lemma 2.2)} \\
&< \infty.
\end{aligned}
$$

On the other hand

$$
W_{1,2} = W_{2,1} = E\left(U_t^2(\boldsymbol{\theta})\left|\frac{\partial}{\partial\alpha_1}g(\boldsymbol{\theta}, F_{t-1})\frac{\partial}{\partial\alpha_2}g(\boldsymbol{\theta}, F_{t-1})\right|\right) \quad = \quad 0
$$

and

$$
\begin{aligned}
W_{1,3} = W_{3,1} = E\left(U_t^2(\boldsymbol{\theta})\left|\frac{\partial}{\partial\alpha_1}g(\boldsymbol{\theta}, F_{t-1})\frac{\partial}{\partial\lambda}g(\boldsymbol{\theta}, F_{t-1})\right|\right) \\
&= E\left(U_t^2(\boldsymbol{\theta})X_{t-1}I_{t-1,1}\right) \\
&= E\left(X_{t-1}I_{t-1,1}V\left(X_t|X_{t-1}\right)\right) \\
&= E\left(X_{t-1}I_{t-1,1}(\alpha_1(1-\alpha_1)X_{t-1}I_{t-1,1} + \alpha_2(1-\alpha_2)X_{t-1}I_{t-1,1} + \lambda)\right) \\
&= \alpha_1(1-\alpha_1)E(X_{t-1}^2I_{t-1,1}) + \lambda E(X_{t-1}I_{t-1,1}) \\
&= p\alpha_1(1-\alpha_1)E(X_{t-1}^2|X_{t-1} \leq R) + p\lambda E(X_{t-1}|X_{t-1} \leq R) \\
&< \infty \text{ (by Remark 2.2);}
\end{aligned}
$$

$$
\begin{aligned}
W_{2,3} = W_{3,2} = E\left(U_t^2(\boldsymbol{\theta})\left|\frac{\partial}{\partial\alpha_2}g(\boldsymbol{\theta}, F_{t-1})\frac{\partial}{\partial\lambda}g(\boldsymbol{\theta}, F_{t-1})\right|\right) \\
&= E\left(U_t^2(\boldsymbol{\theta})X_{t-1}I_{t-1,2}\right) \\
&= E\left(X_{t-1}I_{t-1,2}V\left(X_t|X_{t-1}\right)\right) \\
&= E\left(X_{t-1}I_{t-1,2}(\alpha_1(1-\alpha_1)X_{t-1}I_{t-1,1} + \alpha_2(1-\alpha_2)X_{t-1}I_{t-1,1} + \lambda)\right) \\
&= \alpha_2(1-\alpha_2)E(X_{t-1}^2I_{t-1,2}) + \lambda E(X_{t-1}I_{t-1,2}) \\
&= (1-p)\alpha_2(1-\alpha_2)E(X_{t-1}^2|X_{t-1} > R) + (1-p)\lambda E(X_{t-1}|X_{t-1} > R) \\
&< \infty \text{ (by Remark 2.2).}
\end{aligned}
$$

Therefore condition **(B)** is also satisfied. Finally, note that the determinant of V is

$$\begin{aligned}|V| &= p(1-p)\left(m_2^{(1)}m_2^{(2)} - pu_1^2 m_2^{(2)} - (1-p)u_2^2 m_2^{(1)}\right)\\ &= p(1-p)\left(\sigma_1^2\sigma_2^2 + p\sigma_1^2 u_2^2 + (1-p)u_1^2\sigma_2^2\right) > 0,\end{aligned}$$

which lead us to conclude that V is invertible. Thus condition **(C)** is thereby satisfied. Finally by Theorem 3.2 of Klimko and Nelson (1978) the CLS-estimators $\hat{\boldsymbol{\theta}}_{CLS}$ are asymptotically normal with

$$V = \begin{bmatrix} pm_2^{(1)} & 0 & pu_1 \\ 0 & (1-p)m_2^{(2)} & (1-p)u_2 \\ pu_1 & (1-p)u_2 & 1 \end{bmatrix}$$

and

$$W = \begin{bmatrix} p[\alpha_1(1-\alpha_1)m_3^{(1)} + \lambda m_2^{(1)}] & 0 & p[\alpha_1(1-\alpha_1)m_2^{(1)} + \lambda u_1] \\ 0 & (1-p)[\alpha_2(1-\alpha_2)m_3^{(2)} + \lambda m_2^{(2)}] & (1-p)[\alpha_2(1-\alpha_2)m_2^{(2)} + \lambda u_2] \\ p[\alpha_1(1-\alpha_1)m_2^{(1)} + \lambda u_1] & (1-p)[\alpha_2(1-\alpha_2)m_2^{(2)} + \lambda u_2] & u \end{bmatrix}.$$

□

3.2 Conditional Maximum Likelihood Estimation (CML)

For a fixed value of x_0 the conditional likelihood function for the SETINAR$(2,1)$ model can be shown to be

$$\begin{aligned} L(\boldsymbol{\theta}) &:= P\left(X_1 = x_1, \ldots, X_n = x_n | x_0\right) \\ &= \prod_{t=1}^{n} P(X_t = x_t | X_{t-1} = x_{t-1}) \\ &= \prod_{t=1}^{n} p(x_{t-1}, x_t, \alpha_1 I_{t-1,1} + \alpha_2 I_{t-1,2}, \lambda). \end{aligned}$$

The CML-estimators $\hat{\boldsymbol{\theta}}_{CML} := (\hat{\alpha}_{1,CML}, \hat{\alpha}_{2,CML}, \hat{\lambda}_{CML})$ are obtained maximizing the conditional log-likelihood function

$$l(\boldsymbol{\theta}) := \sum_{t=1}^{n} \log(p(x_{t-1}, x_t, \alpha_1 I_{t-1,1} + \alpha_2 I_{t-1,2}, \lambda).$$

From the partial derivatives of first order we obtain the system

$$
\begin{cases}
\frac{1}{\alpha_i(1-\alpha_i)}\sum_{t=1}^{n} I_{t-1,i}(x_t-\alpha_i x_{t-1}) - \lambda\frac{p(x_{t-1},x_t-1,\alpha_i,\lambda)}{p(x_{t-1},x_t,\alpha_i,\lambda)}I_{t-1,i} & = \; 0, \quad i=1,2 \\
\sum_{t=1}^{n}\left(\frac{p(x_{t-1},x_t-1,\alpha_1,\lambda)}{p(x_{t-1},x_t,\alpha_1,\lambda)}I_{t-1,1} + \frac{p(x_{t-1},x_t-1,\alpha_2,\lambda)}{p(x_{t-1},x_t,\alpha_2,\lambda)}I_{t-1,2}\right) - n & = \; 0
\end{cases}
. \tag{11}
$$

Analytical estimates for this system cannot be found. Thus to solve this system numerical procedures have to be employed. The following results establish consistency and the asymptotic distribution of the CLS-estimators.

Theorem 3.2. *Let $\{X_t\}$ be a SETINAR(2,1) process satisfying (C1)-(C6). Then, there exists a consistent solution $\hat{\boldsymbol{\theta}}_{CML}$ of* (11) *which is a local maximum of $l(\boldsymbol{\theta})$ with probability going to one. Moreover, any other consistent solution of* (11) *coincides with $\hat{\boldsymbol{\theta}}_{CML}$ with probability going to one, when n tends to infinity.*

Theorem 3.3. *Under the assumptions of the Theorem 3.2 and for a fixed value of R the CML-estimators are asymptotically normal, i.e.*

$$
\sqrt{n}(\hat{\boldsymbol{\theta}}_{CML}-\boldsymbol{\theta}) = \sqrt{n}\begin{bmatrix} \hat{\alpha}_{1,CML}-\alpha_1 \\ \hat{\alpha}_{2,CML}-\alpha_2 \\ \hat{\lambda}_{CML}-\lambda \end{bmatrix} \overset{d}{\to} N(0, I(\boldsymbol{\theta})^{-1}), \tag{12}
$$

where $I(\boldsymbol{\theta})$ is the Fisher information matrix.

Proof. *of Theorems 3.2 and 3.3*

In order to find large sample distribution of the CLM-estimators, we will use the same arguments as in Franke and Seligmann (1993, pp. 324-5). The consistency and the asymptotic distribution of the CLS-estimators for the INAR(1) process can be obtained by means of Theorems 2.1 and 2.2 in Billingsley (1961, pp. 10-13). For completeness and reader's convenience Conditions 1.1 and 1.2 of Theorems 2.1 and 2.2 in Billingsley (1961) are given below:

Let $\boldsymbol{\theta} := (\theta_1, \ldots, \theta_r)$ be a parameter which ranges over an open subset Θ of r-dimensional Euclidian space.

(A) For any ξ, the set of η for which $f(\xi, \eta; \boldsymbol{\theta}) > 0$ does not depend on $\boldsymbol{\theta}$. For any ξ and η, $f_u(\xi, \eta; \boldsymbol{\theta})$, $f_{uv}(\xi, \eta; \boldsymbol{\theta})$ and $f_{uvw}(\xi, \eta; \boldsymbol{\theta})$ exist and are continuous throughout $\boldsymbol{\theta}$. (Then for any ξ, $g(\xi, \eta; \boldsymbol{\theta}) = \log f(\xi, \eta; \boldsymbol{\theta})$ is well defined except on a set of $p(\xi, \cdot; \boldsymbol{\theta})$-measure 0, and $g_u(\xi, \eta; \boldsymbol{\theta})$, $g_{uv}(\xi, \eta; \boldsymbol{\theta})$ and $g_{uvw}(\xi, \eta; \boldsymbol{\theta})$ exist and are continuous in $\boldsymbol{\theta}$). For any $\boldsymbol{\theta} \in \Theta$ there exists a neighborhood N of $\boldsymbol{\theta}$ such that for any u, v, w, ξ,

$$\int_X \sup_{\boldsymbol{\theta}' \in N} |f_u(\xi, \eta; \boldsymbol{\theta}')| \lambda(d\eta) < \infty;$$

$$\int_X \sup_{\boldsymbol{\theta}' \in N} |f_{uv}(\xi, \eta; \boldsymbol{\theta}')| \lambda(d\eta) < \infty;$$

$$E_{\boldsymbol{\theta}}(\sup_{\boldsymbol{\theta}' \in N} |g_{uvw}(x_1, x_2; \boldsymbol{\theta}')|) < \infty.$$

Finally, for $u = 1, \ldots, r$

$$E_{\boldsymbol{\theta}}(|g_u(x_1, x_2; \boldsymbol{\theta})|^2) < \infty$$

and if $\sigma_{uv}(\boldsymbol{\theta})$ is defined by

$$\sigma_{uv}(\theta) = E_{\boldsymbol{\theta}}(g_u(x_1, x_2; \boldsymbol{\theta}) g_v(x_1, x_2; \boldsymbol{\theta}))$$

then the $r \times r$ matrix $\sigma(\boldsymbol{\theta}) = (\sigma_{uv})$ is nonsingular.

(B) (i) For each $\theta \in \Theta$, the stationary distribution, which by assumption exists and is unique, has the property that each $\xi \in X$, $p_{\boldsymbol{\theta}}(\xi, \cdot)$ is absolutely continuous with respect to $p_{\boldsymbol{\theta}}(\cdot)$:

$$p_{\boldsymbol{\theta}}(\xi, \cdot) \ll p_{\boldsymbol{\theta}}(\cdot).$$

(ii) There is some $\delta > 0$ (which may depend on $\boldsymbol{\theta}$) such that for $u = 1, \ldots, r$,

$$E_{\boldsymbol{\theta}}(|g_u(x_1 x_2; \boldsymbol{\theta})|^{2+\delta}) < \infty.$$

Note that in the context of the SETINAR(2;1) model conditions (A) and (B) have to be adapted.

The first partial derivatives of the transition function are given by

$$\frac{\partial p(x_{t-1},x_t)}{\partial\lambda} = I_{t-1,1}\frac{\partial p(x_{t-1},x_t,\alpha_1,\lambda)}{\partial\lambda}+I_{t-1,2}\frac{\partial p(x_{t-1},x_t,\alpha_2,\lambda)}{\partial\lambda} \tag{13}$$

and for $i=1,2$,

$$\frac{\partial p(x_{t-1},x_t)}{\partial\alpha_i} = I_{t-1,i}\frac{x_{t-1}}{1-\alpha_i}\{p(x_{t-1}-1,x_t-1,\alpha_i,\lambda)-p(x_{t-1},x_t,\alpha_i,\lambda)\}. \tag{14}$$

From expressions (13) and (14) it follows easily that the first derivatives of the logarithm of the transition function are

$$\frac{\partial \log p(x_{t-1},x_t)}{\partial\lambda} = \sum_{i=1}^{2} I_{t-1,i}\frac{\partial}{\partial\lambda}\log p(x_{t-1},x_t,\alpha_i,\lambda) \tag{15}$$

$$\frac{\partial \log p(x_{t-1},x_t)}{\partial\alpha_i} = I_{t-1,i}\frac{\partial}{\partial\alpha_i}\log p(x_{t-1},x_t,\alpha_i,\lambda),\ i=1,2. \tag{16}$$

Equations (13)-(16) allow us to conclude that each regime falls into the INAR structure considered by Franke and Seligmann (1993). These authors showed that for the Poisson distribution, as the distribution of innovations, the following set of conditions hold:

(C1) The set $\{k : P(Z_t=k)=f(k,\lambda)=e^{-\lambda}\frac{\lambda^k}{k!}>0\}$ does not depend of λ;

(C2) $E[Z_t^3]=\lambda^3+3\lambda^2+\lambda<\infty$;

(C3) $P(Z_t=k)$ is three times continuously differentiable with respect to λ;

(C4) For any $\lambda'\in B$, where B is an open subset of $\mathbb{R}$, there exists a neighborhood U of λ' such as

1. $\displaystyle\sum_{k=0}^{\infty}\sup_{\lambda\in U} f(k,\lambda)<\infty$;
2. $\displaystyle\sum_{k=0}^{\infty}\sup_{\lambda\in U}\left|\frac{\partial f(k,\lambda)}{\partial\lambda_i}\right|<\infty,\ \ i=1,\ldots,n$;
3. $\displaystyle\sum_{k=0}^{\infty}\sup_{\lambda\in U}\left|\frac{\partial^2 f(k,\lambda)}{\partial\lambda_i\partial\lambda_j}\right|<\infty,\ \ i,j=1,\ldots,n$.

Moreover for the SETINAR(2;1) model it is necessary to verify the following conditions (analogous to conditions (C5) and (C6) in Franke and Seligmann (1993)):

(C5) For any $\lambda' \in B$ there exists a neighborhood U of λ' and the sequences $\psi_1(n) = const1.n$, $\psi_{11}(n) = const2.n^2$, and $\psi_{111}(n) = const3.n^3$, with $const1, const2, const3$ suitable constants, and $n \geq 0$ such as $\forall \lambda \in B$ e $\forall m \leq n$, with $P(Z_t)$ non-vanishing,

$$\begin{aligned}
\left|\frac{\partial f(m,\lambda)}{\partial \lambda}\right| &\leq \psi_1(n) f(m,\lambda);\\
\left|\frac{\partial^2 f(m,\lambda)}{\partial \lambda^2}\right| &\leq \psi_{11}(n) f(m,\lambda);\\
\left|\frac{\partial^3 f(m,\lambda)}{\partial \lambda^3}\right| &\leq \psi_{111}(n) f(m,\lambda)
\end{aligned}$$

and with respect to the stationary distribution of the process (X_t)

$$\begin{aligned}
E[\psi_1^3(X_1)] &< \infty;\\
E[X_1 \psi_{11}(X_2)] &< \infty;\\
E[\psi_i(X_1)\psi_{11}(X_2)] &< \infty;\\
E[\psi_{111}(X_1)] &< \infty.
\end{aligned}$$

(C6) The Fisher information matrix $I(\boldsymbol{\theta})$ is non-singular, which guarantees that the parameters of SETINAR(2;1) model are not redundant.

The first set of conditions which are related with the innovations distributions were proved by Franke and Seligmann (1993) and the set of conditions in **(C5)** related with the stationary distribution of the SETINAR$(2,1)$ model follows by Lemmas 2.1 and 2.2. To prove condition **(C6)**, the determinant of the Fisher information matrix is given by

$$\begin{aligned}
|I(\boldsymbol{\theta})| &= \sum_{i=1}^{2} P(I_{X_1,3-i}=1)^2 P(I_{X_1,i}=1) E[(\frac{\partial}{\partial \alpha_i}\log p(X_1,X_2,\alpha_i,\lambda))^2 | I_{X_1,i}=1] \times |A_{3-i}|,\\
&= p(1-p)\sum_{i=1}^{2} P(I_{X_1,3-i}=1) E[(\frac{\partial}{\partial \alpha_i}\log p(X_1,X_2,\alpha_i,\lambda))^2 | I_{X_1,i}=1] \times |A_{3-i}|,
\end{aligned}$$

where, for $i = 1,2$,

$$\begin{aligned}
(A_i)_{11} &= (E[(\frac{\partial}{\partial \alpha_i}\log p(X_1,X_2,\alpha_i,\lambda))^2 | I_{X_1,i}=1];\\
(A_i)_{12} &= (A_i)_{21} = E\left[\frac{\partial}{\partial \alpha_i}\log p(X_1,X_2,\alpha_i,\lambda)\frac{\partial}{\partial \lambda}\log p(X_1,X_2,\alpha_i,\lambda) | I_{X_1,i}=1\right];\\
(A_i)_{22} &= E[(\frac{\partial}{\partial \lambda}\log p(X_1,X_2,\alpha_i,\lambda))^2 | I_{X_1,i}=1].
\end{aligned}$$

It is important to stress that matrices A_i, $i = 1,2$, has the same structure as the Fisher information matrix analyzed by Franke and Seligmann (1993), which implies that the same arguments can be used to prove that A_i has positive determinant. The matrix A_i, $i = 1,2$, is, e.g., non-singular if the matrix with entries

$$\sum_{m_i}(\alpha_i,\lambda) = \begin{bmatrix} E\left(\left(\frac{\partial}{\partial\alpha_i}\log p(X_1,X_2,\alpha_i,\lambda)\right)^2 \middle| X_1 = m_i\right) & E\left(\begin{array}{c}\frac{\partial}{\partial\alpha_i}\log p(X_1,X_2,\alpha_i,\lambda)\times \\ \frac{\partial}{\partial\lambda}\log p(X_1,X_2,\alpha_i,\lambda)\end{array}\middle| X_1 = m_i\right) \\ E\left(\begin{array}{c}\frac{\partial}{\partial\alpha_i}\log p(X_1,X_2,\alpha_i,\lambda)\times \\ \frac{\partial}{\partial\lambda}\log p(X_1,X_2,\alpha_i,\lambda)\end{array}\middle| X_1 = m_i\right) & E\left(\left(\frac{\partial}{\partial\lambda}\log p(X_1,X_2,\alpha_i,\lambda)\right)^2 \middle| X_1 = m_i\right) \end{bmatrix}$$

is non-singular for a set of m_i, $m_1 \leq R$ and $m_2 > R$, with positive measure under the stationary distribution. Franke and Seligmann (1993) proved that

$$\begin{aligned} p(m,n,\alpha_i,\lambda) &= \alpha_i p(m-1,n-1,\alpha_i,\lambda) + (1-\alpha_i)p(m-1,n,\alpha_i,\lambda), \\ \frac{\partial p(m,n,\alpha_i,\lambda)}{\partial\alpha_i} &= \frac{m}{1-\alpha_i}[p(m-1,n-1,\alpha_i,\lambda) - p(m,n,\alpha_i,\lambda)] \\ \frac{\partial p(m,n,\alpha_i,\lambda)}{\partial\lambda} &= \left(\frac{n}{\lambda}-1\right)p(m,n,\alpha_i,\lambda) - \left(\frac{m\alpha_i}{\lambda}\right)p(m-1,n-1,\alpha_i,\lambda), \end{aligned}$$

and considering $D(m,n,\alpha_i) = \dfrac{p(m-1,n-1,\alpha_i,\lambda)}{p(m,n,\alpha_i,\lambda)}$ we have

$$\begin{aligned} \frac{\lambda^2(1-\alpha_i)^2}{m_i^2} det\sum_{m_i} &= Var\left[(D(m_i,X_2,\alpha_i)-1)(X_2-\lambda-m_i\alpha_i D(m_i,X_2,\alpha_i))\right] - \\ &\quad -Cov((D(m_i,X_2,\alpha_i)-1)^2,(X_2-\lambda-m_i\alpha_i D(m_i,X_2,\alpha_i))^2), \end{aligned}$$

which is positive for a set of m_i's such that $m_1 \leq R$ and $m_2 > R$, with positive measure under the stationary distribution, and therefore condition **(C6)** is also satisfied.

Since each regime of the SETINAR(2;1) model falls, in term of derivatives of log-likelihood, into the INAR structure considered by Franke and Seligmann (1993), and according with these authors conditions **(C1)-(C6)** imply the conditions **(A)** and **(B)** of Theorems 2.1. and 2.2 in Billingsley (1961) and thus the results in Theorems 3.2 and 3.3 are also valid for the SETINAR(2,1) process. □

4 Simulation study

The aim of this section is to illustrate the theoretical findings given in Section 3 and to assess the small, moderate and large sample behavior of the CLS- and CML-estimators. The simulation study contemplates the following combination of α's and λ: $\alpha_1 = \{0.2, 0.8\}$, $\alpha_2 = \{0.1, 0.65\}$ and $\lambda = \{3, 7\}$. For each combination of these parameters, the value of R was chosen such that at least 50% of the observations are in the first regime. Hence we consider eight distinct SETINAR$(2, 1)$ models with Poisson innovations; see Table 1.

(Table 1 about here)

For each model, time series of length $n = 50, 100, 200, 500$ with 1000 independent replicates were simulated. The results are summarized in Tables 2 and 3.

(Table 2 about here)

(Table 3 about here)

A closer look at the tables shows the superiority of the CML method in terms of both bias and mean square error (MSE), with special relevance for small and moderate samples. Figures 3 and 4 display the boxplots of CLS and CML estimates for each model.

(Figure 3 about here)

(Figure 4 about here)

Note that for $\alpha_1 = 0.2$ the biases are more scattered than for $\alpha_1 = 0.8$, regardless the value of λ. As expected, both the bias and the skewness are also reduced when the sample size increases. This is in agreement with the asymptotic properties of the estimators: unbiasedness and consistency. Moreover, as larger the difference between α_1 and α_2, in absolute value, the better the performance of both methods. This is in contrast with the case of values of α_1 and α_2 being too close from each other, since in this case it seems more difficult for both the CLS and the CML methods to distinguish between the two regimes.

5 Conclusions

This paper has introduced a class of self-exciting threshold integer-valued autoregressive models driven by independent Poisson-distributed random variables. The stationarity and ergodicity of the process are established. Least squares-type and likelihood-based estimators of the model parameters were derived and their asymptotic properties obtained. Potential issues of future research include to extend the results for general SETINAR $(k; p_1, \ldots, p_k; q_1, \ldots, q_k)$ models including an arbitrary number of threshold as well as autoregressive and moving average parameters. This remains a topic of future research.

References

[1] Ahn, S., Gyemin, L., Jongwoo, J. (2000). Analysis of the M/D/1-type queue based on an integer-valued autoregressive process. *Oper. Res. Lett.* 27:235-241.

[2] Alosh, M. (2009). The impact of missing data in a generalized integer-valued autoregression model for count data. *J. Biopharm. Statist.* 19:1039-1054.

[3] Amaral, L.F., Souza, R.C., Stevenson, M. (2008). A smooth transition periodic autoregressive (STPAR) model for short-term load forecasting. *Int. J. Forecast.* 24:603–615.

[4] Billingsley, P. (1961). *Statistical Inference for Markov Processes.* Chicago: University of Chicago Press.

[5] Billingsley, P. (1979). *Probability and Measure.* John Wiley & Sons, Inc.

[6] Boero, G., Marrocu, E. (2004). The performance of SETAR models: a regime conditional evaluation of point, interval and density forecasts. *Int. J. Forecast.* 20:305–320.

[7] Brännäs, K. and Quoreshi, A. M. M. S. (2010). Integer-valued moving average modelling of the number of transactions in stocks. *Appl. Finan. Econ.* 20:1429–1440.

[8] Doukhan, P., Latour, A., Oraichi, D. (2006). Simple integer-valued bilinear time series model. *Adv. Appl. Probab.* 38:559–578.

[9] Drost, F. C., van den Akker, R., Werker, B. (2008). Note on integer-valued bilinear time series models. *Statist. Probab. Lett.* 78:992-996.

[10] Fu, Q., Fu, H., Sun, Y. (2004). Self-exciting threshold auto-regressive model (SETAR) to forecast the well irrigation rice water requirement. *Nature and Sciences* 2:36–43.

[11] Franke, J., Seligmann, T. (1993). Conditional maximum likelihood estimates for INAR(1) processes and their application to modelling epileptic seizure counts. In: Subba Rao, T., ed. *Developments in Time Series Analysis.* London: Chapman & Hall, pp. 310-330.

[12] Hall, A. Scotto, M. G., Cruz, J. P. (2010). Extremes of integer-valued moving average sequences. *Test* 19:359–374.

[13] Jung, R. C., Tremayne, A. R. (2006). Binomial thinning models for integer time series. *Stat. Model.* 6:81-96.

[14] Jung, R. C. and Tremayne, A. R. (2010). Useful models for time series of counts or simply wrong ones? *Adv. Stat. Anal.* (to appear).

[15] Klimko, L. A., Nelson, P. I. (1978). On conditional least squares estimation for stochastic process. *Ann. Statist.* 6:629–642.

[16] Lambert, D., Liu, C. (2006). Adaptive thresholds: monitoring streams of network counts. *J. Am. Stat. Assoc.* 101:78-88.

[17] McCabe, B. P. M., Martin, G. M. (2005). Bayesian prediction of low count time series. *Int. J. Forecast.* 21:315-330.

[18] McKenzie, E., 2003. Discrete variate time series. In: Shanbhag, D. N. Rao, C. R., eds. *Handbook of Statistics.* Amsterdam: Elsevier, pp. 573-606.

[19] Monteiro, M., Pereira, I., Scotto, M. G. (2008). Optimal alarm systems for count processes. *Comm. Statist. Theor. Meth.* 37:3054-3076.

[20] Pai, J., Pedersen, H. W. (1999). Threshold models of the term structura of interest rate. *Proceedings of the 9th International AFIR Colloquium.* 387–400.

[21] Potter, S. M. (1995). A nonlinear approach to U.S. GNP. *J. Appl. Econ.* 10:109–125.

[22] Scotto, M. G., Guedes Soares, C. (2000). Modelling the long-term time series of significant wave height with non-linear threshold models. *Coast. Eng.* 40, 313–327.

[23] Steutel, F., van Harn, K. (1979). Discrete analogues of self-decomposability and stability. *Ann. Probab.* 7, 893–899.

[24] Thyregod, P., Carstensen, J., Madsen, H., Arnbjerg-Nielsen, K. (1999). Integer valued autoregressive models for tipping bucket rainfall measurements. *Environmetrics* 10:395–411.

[25] Tong, H. (1977). Some comments on the Canadian lynx data. *J. Roy. Statist. Soc. Ser. A* 140:432–435.

[26] Tong, H. (1983). *Threshold Models in Nonlinear Time Series Analysis*. London: Springer.

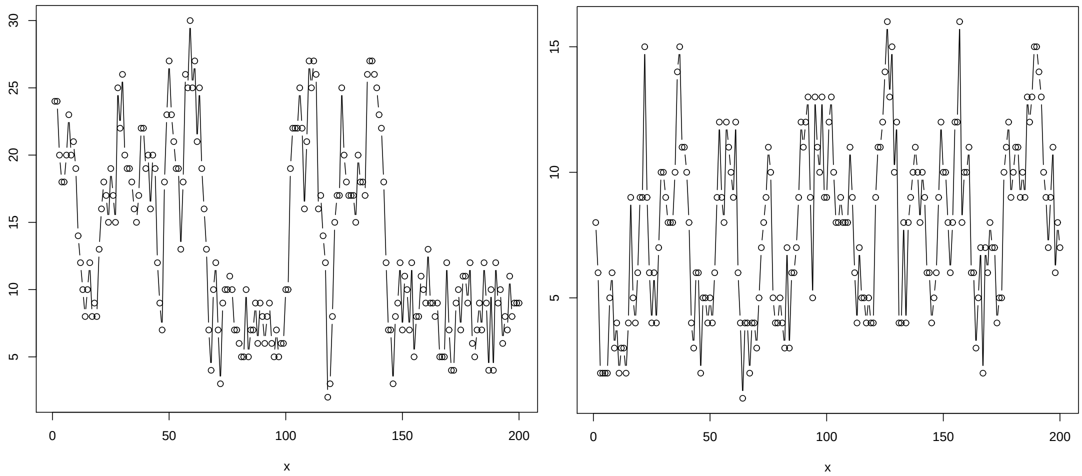
[27] Tong, H. (1990). *Non-linear Time Series: a Dynamical System Approach.* Oxford: Clarendon Press.

[28] Weiß, C. H. (2008a). The combined INAR(p) models for time series of counts. *Statist. Probab. Lett.* 78:1817-1822.

[29] Weiß, C. H. (2008b). Thinning operations for modelling time series of counts-a survey. *Adv. Statist. Anal.* 92:319-341.

[30] Weiß, C. H. (2008c). Controlling jumps in correlated processes of Poisson counts. *Appl. Stochastic Models Bus. Ind.* 25:551-564.

[31] Zhou, J., Basawa, I. V. (2005). Least-squared estimation for bifurcation autoregressive processes. *Statist. Probab. Lett.* 74:77-88.



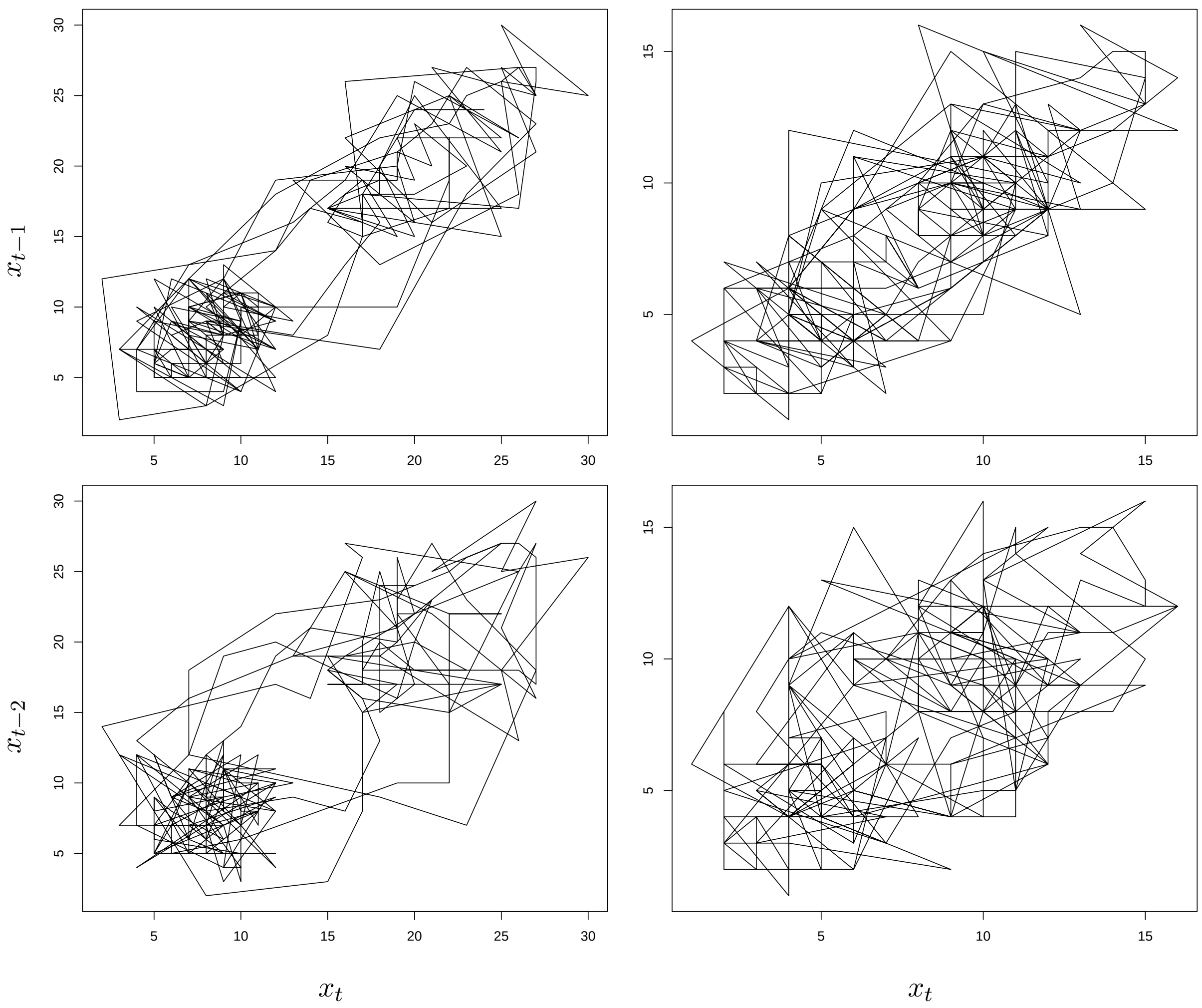
(a) $\lambda = 7;\ \alpha_1 = 0.2;\ \alpha_2 = 0.65;\ R = 14$

(b) $\lambda = 3;\ \alpha_1 = 0.5;\ \alpha_2 = 0.65;\ R = 7$

Figure 1: Simulated sample paths from the SETINAR(2,1) model.

	$\lambda = 3$				$\lambda = 7$		
	α_1	α_2	R		α_1	α_2	R
A1	0.2	0.1	4	B1	0.2	0.1	8
A2	0.2	0.65	6	B2	0.2	0.65	14
A3	0.8	0.1	9	B3	0.8	0.1	21
A4	0.8	0.65	11	B4	0.8	0.65	27

Table 1: Parameters of the SETINAR$(2, 1)$ models



(a) $\lambda = 7;\ \alpha_1 = 0.2;\ \alpha_2 = 0.65;\ R = 14$ (b)$\lambda = 3;\ \alpha_1 = 0.5;\ \alpha_2 = 0.65;\ R = 7$

Figure 2: Directed scatter diagrams of the realizations in Figure 1.

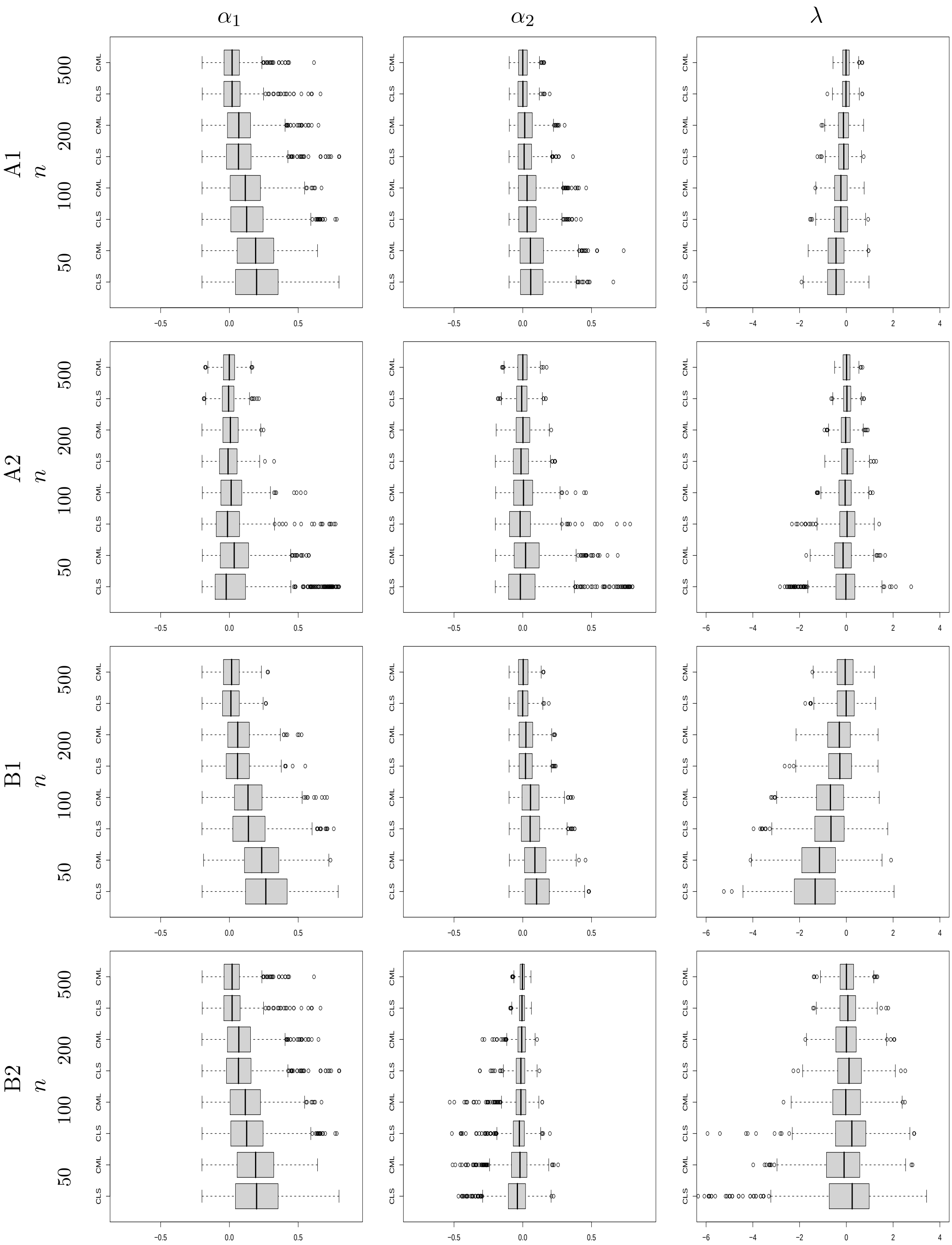


Figure 3: Boxplots of the biases for $\boldsymbol{\theta}$ models A1, A2, B1 and B2, with respectively $\boldsymbol{\theta} = (0.2, 0.1, 3)$, $\boldsymbol{\theta} = (0.2, 0.65, 3)$, $\boldsymbol{\theta} = (0.2, 0.1, 7)$ and $\boldsymbol{\theta} = (0.2, 0.65, 7)$.

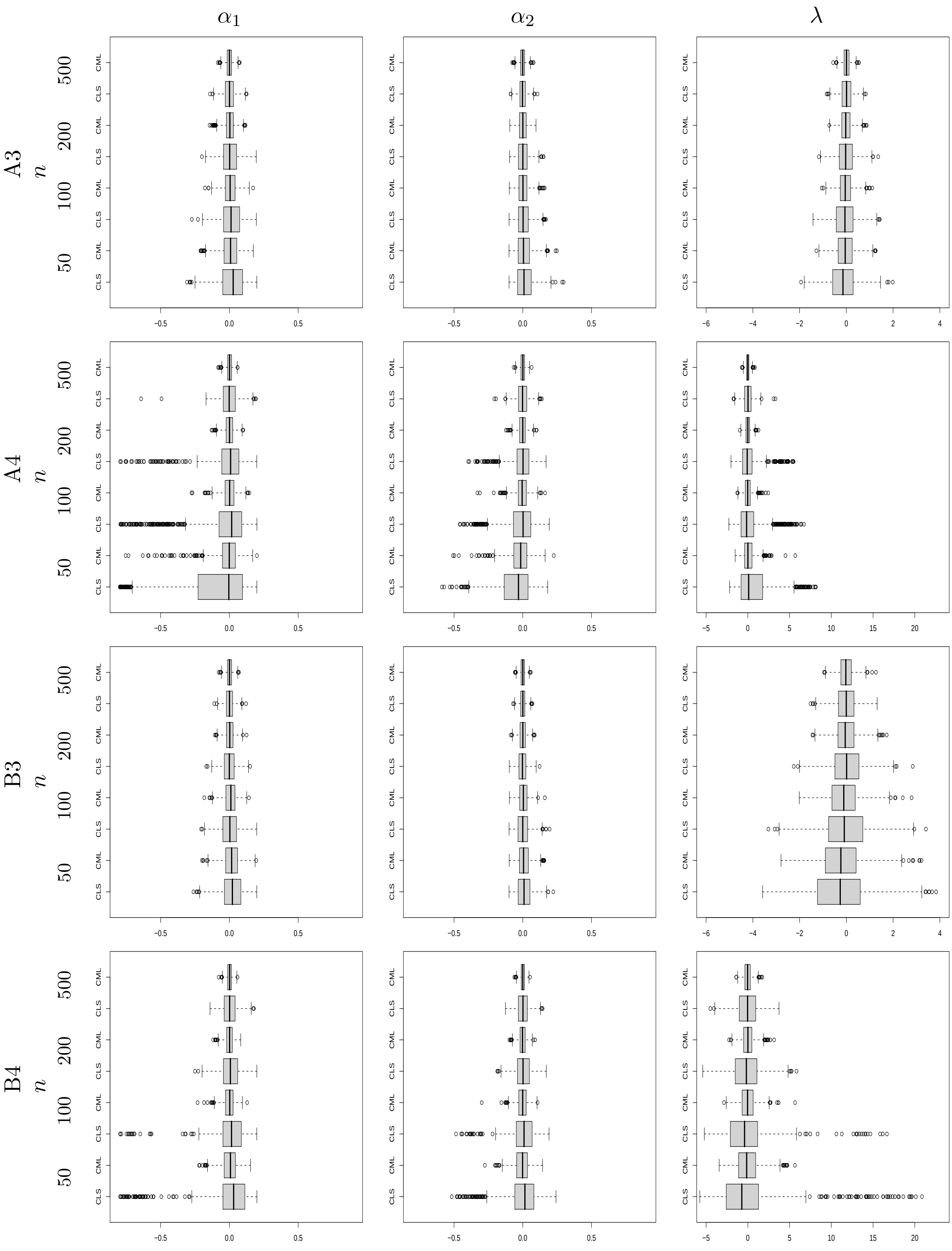


Figure 4: Boxplots of the biases for $\boldsymbol{\theta}$ models A3, A4, B3 and B4, with respectively $\boldsymbol{\theta} = (0.8, 0.1, 3)$, $\boldsymbol{\theta} = (0.8, 0.65, 3)$, $\boldsymbol{\theta} = (0.8, 0.1, 7)$ and $\boldsymbol{\theta} = (0.8, 0.65, 7)$.

Model	Par	N = 50 CLS	50 CML	100 CLS	100 CML	200 CLS	200 CML	500 CLS	500 CML
A1	α_1	0.413 (0.095)	0.388 (0.067)	0.336 (0.051)	0.322 (0.042)	0.280 (0.031)	0.276 (0.025)	0.222 (0.011)	0.221 (0.010)
	α_2	0.176 (0.020)	0.177 (0.021)	0.141 (0.010)	0.141 (0.011)	0.120 (0.005)	0.121 (0.005)	0.101 (0.002)	0.101 (0.002)
	λ	2.556 (0.464)	2.565 (0.422)	2.762 (0.215)	2.765 (0.201)	2.880 (0.113)	2.873 (0.107)	2.991 (0.049)	2.986 (0.045)
	p	0.717		0.717		0.718		0.718	
A2	α_1	0.243 (0.049)	0.249 (0.024)	0.201 (0.018)	0.219 (0.012)	0.193 (0.008)	0.208 (0.006)	0.192 (0.004)	0.198 (0.003)
	α_2	0.585 (0.018)	0.609 (0.014)	0.622 (0.007)	0.635 (0.005)	0.635 (0.003)	0.643 (0.002)	0.643 (0.001)	0.646 (0.001)
	λ	2.905 (0.530)	2.849 (0.288)	3.028 (0.237)	2.946 (0.161)	3.042 (0.123)	2.981 (0.089)	3.038 (0.055)	3.012 (0.040)
	p	0.698		0.673		0.661		0.655	
B1	α_1	0.470 (0.120)	0.428 (0.085)	0.352 (0.052)	0.341 (0.043)	0.263 (0.019)	0.266 (0.018)	0.212 (0.008)	0.215 (0.007)
	α_2	0.212 (0.027)	0.197 (0.020)	0.162 (0.013)	0.160 (0.011)	0.124 (0.005)	0.126 (0.005)	0.102 (0.003)	0.104 (0.002)
	λ	5.630 (3.431)	5.831 (2.462)	6.252 (1.442)	6.293 (1.197)	6.700 (0.579)	6.677 (0.563)	6.965 (0.286)	6.943 (0.254)
	p	0.567	0.567	0.568	0.568	0.569	0.569	0.569	0.569
B2	α_1	0.219 (0.033)	0.237 (0.020)	0.191 (0.014)	0.209 (0.010)	0.186 (0.007)	0.200 (0.006)	0.192 (0.003)	0.198 (0.002)
	α_2	0.599 (0.014)	0.616 (0.011)	0.617 (0.007)	0.628 (0.006)	0.632 (0.003)	0.640 (0.002)	0.643 (0.001)	0.646 (0.001)
	λ	7.025 (2.116)	6.821 (1.241)	7.176 (1.079)	6.995 (0.738)	7.142 (0.575)	7.011 (0.426)	7.074 (0.249)	7.024 (0.181)
	p	0.673	0.673	0.684	0.684	0.688	0.688	0.688	0.688

Table 2: Sample mean and mean square error (in brackets) for models A1, A2, B1 and B2, with $\boldsymbol{\theta} = \{(0.2, 0.1, 3), (0.2, 0.65, 3), (0.2, 0.1, 7), (0.2, 0.65, 7)\}$, respectively.

Model	Par	N 50		100		200		500	
		CLS	CML	CLS	CML	CLS	CML	CLS	CML
	α_1	0.820	0.804	0.813	0.805	0.805	0.803	0.800	0.799
		(0.010)	(0.005)	(0.007)	(0.003)	(0.005)	(0.002)	(0.002)	(0.001)
	α_2	0.117	0.110	0.106	0.103	0.101	0.101	0.099	0.099
		(0.005)	(0.004)	(0.003)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
A3	λ	2.869	2.953	2.933	2.974	2.977	2.985	3.005	3.010
		(0.388)	(0.196)	(0.264)	(0.116)	(0.175)	(0.067)	(0.074)	(0.027)
	p	0.796		0.796		0.795		0.795	
	α_1	0.804	0.784	0.813	0.799	0.807	0.801	0.799	0.802
		(0.014)	(0.006)	(0.010)	(0.002)	(0.006)	(0.001)	(0.003)	(4.9e-4)
	α_2	0.396	0.387	0.403	0.396	0.402	0.400	0.398	0.400
		(0.010)	(0.006)	(0.006)	(0.003)	(0.003)	(0.001)	(0.001)	(0.001)
A4	λ	2.941	3.079	2.900	2.997	2.949	2.987	3.014	2.993
		(0.878)	(0.383)	(0.592)	(0.164)	(0.363)	(0.089)	(0.172)	(0.035)
	p	0.749		0.747		0.747		0.745	
	α_1	0.821	0.814	0.804	0.807	0.799	0.801	0.800	0.801
		(0.008)	(0.004)	(0.006)	(0.002)	(0.003)	(0.001)	(0.001)	(4.9e-04)
	α_2	0.112	0.109	0.103	0.105	0.098	0.100	0.100	0.100
		(0.004)	(0.002)	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)	(3.4e-04)
B3	λ	6.733	6.808	6.947	6.897	7.025	6.986	6.998	6.988
		(1.684)	(0.935)	(1.213)	(0.549)	(0.549)	(0.269)	(0.233)	(0.114)
	p	0.783	0.783	0.783	0.783	0.783	0.783	0.783	0.783
	α_1	0.797	0.802	0.807	0.799	0.809	0.799	0.801	0.801
		(0.033)	(0.004)	(0.016)	(0.002)	(0.006)	(0.001)	(0.003)	(3.8e-04)
	α_2	0.645	0.643	0.655	0.647	0.655	0.648	0.650	0.650
		(0.015)	(0.004)	(0.008)	(0.002)	(0.004)	(0.001)	(0.002)	(3.1e-04)
B4	λ	7.017	6.988	6.765	7.023	6.782	7.019	6.975	6.984
		(17.078)	(2.241)	(8.067)	(1.047)	(3.570)	(0.581)	(2.003)	(0.234)
	p	0.654	0.654	0.657	0.657	0.657	0.657	0.656	0.656

Table 3: Sample mean and mean square error (in brackets) for models A3, A4, B3 and B4, with $\boldsymbol{\theta} = \{(0.8, 0.1, 3), (0.8, 0.65, 3), (0.8, 0.1, 7), (0.8, 0.65, 7)\}$, respectively.